DOCUMENT RESUME

ED 443 429 IR 057 845

AUTHOR	Rudner, Lawrence
TITLE	Who Is Going To Mine Digital Library Resources? And How?
PUB DATE	2000-05-00
NOTE	13p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Electronic Libraries; Information Literacy; *Information Scientists; *Information Seeking; Internet; *Online Searching; Professional Personnel; *Search Intermediaries; Search Strategies; Users (Information)
IDENTIFIERS	ERIC; ERIC Clearinghouse on Assessment and Evaluation

ABSTRACT

As use of the Internet grows as a research tool, patrons have become increasingly less dependent on librarians and other expert intermediaries. Examining the quality of on-line searches, this paper argues that researchers and other Internet users do not look for and hence do not find the best resources. For two days in early November 1998, all patrons wanting to search the ERIC database installed at the ERIC Clearinghouse on Assessment and Evaluation (ERIC/AE) Web site were required to complete a 10-item background questionnaire. For each patron, the following information was tracked: maximum number of "OR's" in their searches as measure of search quality; number of queries per session; whether they used the thesaurus or free-text search engine; number of hits examined; and amount of time devoted to searching the ERIC database per session. The paper concludes that ready access to resources can lead to decreased research quality and ill-informed practice. Digital resources must be developed with expert intermediaries and contain pre-selected resources if they are to be of service. (Contains 11 references.) (AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED 443 429

Who Is Going to Mine Digital Library Resources? And How?

Lawrence Rudner
ERIC Clearinghouse on Assessment and Evaluation
Department of Measurement, Statistics, and Evaluation
University of Maryland, College Park
rudner@ericae.net

L. Rudner

Abstract

As use of the Internet grows as a research tool, patrons have become increasingly less dependent on librarians and other expert intermediaries. Examining the quality of on-line searches, the author argues that researchers and other internet users do not look for and hence do not find the best resources. He concludes that ready access to resources can lead to decreased research quality and ill-informed practice. Digital resources must be developed with expert intermediaries and contain pre-selected resources if they are to be a service.

Introduction

Until the mid 1980's, most database searching was conducted by expert intermediaries. Reference librarians familiar with the database and trained in information retrieval would conduct searches for the end-user and then present to the user a highly relevant set of references.

In my experience as an end-user, it was a long process. I would need to set up a reference interview with the reference librarian. A few days later I would then get back from the intermediary 30 to 100 citations that were of potential interest. Sometimes I would identify promising citations and the reference professional would then conduct a search based on those potential pearls. I would take my resulting list of abstracts and go to the library where I would then spend hours looking for the source material. Finally, I would find a few key articles, check the references in those articles, and go back to the library to find them. The process would take weeks and was

IR057845

BEST COPY AVAILABLE

very dependent on the reference interview.

That has changed. Evans (Evans 1995) showed that mediated searching peaked through the mid 1980s, then began a sharp decline, while the average cost per search rose steadily throughout the study period. The advent of the compact disc (CD-ROM) work station, with no user costs and direct user searching, altered the use patterns of the mediated search se ces. While the use of CD-ROM searching skyrocketed, dial-up information services steadily declined. Evans noted that there appeared to be a greater willingness on the part of the end-user to invest time and physical effort, with the possibility of error or omission, rather than spend money for a fast, sure, guaranteed product.

The Internet provides the next leap forward with regard to end-user searching. EBSCOHost, OCLC First Search, CatchWord, JSTOR, Highwire, and ERIC now provide instant access to the full-text of articles. End-users can conduct their own searches, read articles on line and even have those articles instantly ready as they write their paper. They can readily find text that they want to quote, and they can readily examine and reexamine key sections of relevant articles. In some instances, they can even click on cited references and retrieve those articles (Stanford's Highwire has that feature). I must confess that I am not accustomed to this technological capability. In writing one paper, I must have retrieved the same three documents ten times each. But I am certain I will adapt and become more efficient in using these new tools, just as I was able to make the transition from having a secretary type my papers to using a word processor.

It appears that the ready availability of digital libraries will be a boon to research and practice. Researchers will be better able to build on past findings; practitioners will be able to base their actions on information. But this is predicated on the assumption that the end-user will be able to identify relevant, high quality documents. Researchers are supposed to be comprehensive in their examination of the literature; practitioners are supposed to base their actions and policies on the best available information.

Extending the work of Hertzberg and Rudner (Hertzberg and Rudner 1999), this paper presents data that questions that assumption. Noting that the quality of most end-user searching is horrible, the paper examines the implications for information professionals.

Method

For two days in early November 1998, all patrons wanting to search the ERIC database installed at the ERIC/AE website were required to complete

a 10-item background questionnaire. For each patron, we then tracked:

a. the maximum number of OR's in their searches as a measure of search quality,
b. the number of queries per session,
c. whether they used the thesaurus or free-text search engine,
d. number of hits examined, and
e. the amount of time devoted to searching the ERIC database per session.

Data were collected on 4,086 user sessions. Because some browsers were not set to accept identifiers, we were not always able to relate background data to session information. Accordingly, our analysis is based on the 3,420 users with background and corresponding session information.

Participation in the study was entirely voluntary; patrons could go elsewhere to search the ERIC database. However, our questionnaire was short and our data collection was unobtrusive. Based on the prior week's log, we estimate our retention rate was more than 90%.

Results

We asked our end-users, "What is the primary purpose of your search today?". As shown in Table 1, most patrons were searching in preparation of a research report.

Table 1: Purpose of searching the ERIC database

Purpose	N	Percent
Research report preparation	1825	53.4%
Class assignment	601	17.6
Professional interest	554	16.2
Lesson planning	177	5.2
Background for policy making	175	5.1
Classroom management	88	2.6
TOTAL	3240	100.0%

Based on their stated purposes, one would expect a sizable number of end-users to be trying to be comprehensive in their efforts. One would expect a large number of citations to be examined and a fair amount of time to be spent on searching.

As shown in Table 2, however, this was not the case. Users typically looked at 3 - 5 hits and spent about five minutes searching. Researchers, College Professors and K-12 librarians tended to look at the most number of potentially relevant citations and had the largest variation in the number of hits examined, but the averages for all groups are terribly low.

Table 2: Searching Characteristics for Select User Groups

		Hits Examined		Time	
	n	mean	std dev	median	sir
K-12 Adminis.	121	3.15	5.24	414	373
Researcher	445	4.85	10.23	376	408
College Professor	209	5.58	15.09	361	345
K-12 Teacher	641	2.88	4.95	331	347
UG Student	380	2.82	5.11	281	272
Grad Student	896	3.71	8.52	391	362
Parent	72	2.14	3.87	304	350
College Librarian	96	3.11	5.41	207	288
K-12 Librarian	71	6.80	23.71	301	400
All Users	3420	3.65	8.65	352	351

All Users	3420	3.65	8.65	352	351

Most variables were fairly normally distributed. Accordingly, means and standard deviations (std dev) are presented in the table. The amount of time spent searching, however, was quite skewed. Central tendency and variability for time are represented by medians and semi-interquartile ranges (sir).

Five minutes is not much time to spend searching, especially if one is trying to be comprehensive. Conceivably, end-users would not need to spend much time searching if first they compose a good search query. Such a search strategy would quickly find the best and most relevant documents. However, as shown in Table 3, end-user search strategies do not appear to be very good.

To provide a perspective on end-user search strategies, we compared information about the end-user strategies we were tracking to information about the strategies of expert groups.

1. ERIC experts - the search strategies developed by the top reference librarians across the entire ERIC system used in the 84 prepackaged search strategies at the ERIC Clearinghouse on Assessment and Evaluation, and the number of queries used in responding to patron questions by the reference staff at the Clearinghouse, and
2. Experienced searchers - the search strategies, number of queries, and use of the on-line thesaurus by the 33 respondents who indicated that they have extensive experience with the ERIC database.

These expert groups averaged two or three "OR" operators in their query (i.e., 3 or 4 terms) and tended to use the ERIC thesaurus. ERIC experts averaged more than five queries and had a much larger range in the number of queries. In contrast, most patrons used very few ORs, conducted very few queries, and tended not to use the on-line thesaurus.

Table 3: Searching Characteristics for Different User Groups

		Number of ORs		N Queries		Thesaurus Use
	n	mean	std dev	mean	std dev	%
ERIC Experts		2.90	2.80	5.40	4.30	100
Experienced	33	2.37	6.40	2.09	1.89	71.9

College Librarian	96	.91	3.89	2.66	3.26	46.8
K-12 Librarian	71	.10	.42	2.51	2.52	29.6
K-12 Adminis.	121	.36	.92	2.93	2.59	37.1
Researcher	445	.42	1.26	3.04	3.69	37.6
College Professor	209	.37	1.10	2.49	2.46	44.6
K-12 Teacher	641	.42	1.52	2.63	2.66	37.3
UG Student	380	.39	1.99	2.85	2.89	24.7
Grad Student	896	.51	2.06	2.75	2.66	44.0
Parent	72	.32	1.11	2.44	3.27	38.6
All Users	3420	.44	1.77	2.75	2.95	38.7

Discussion

To partially answer the questions raised in the title of this paper -- "Who is going to mine digital library resources? And how?" -- today's end-users are not capable of mining today's digital libraries, let alone the more comprehensive digital libraries of the foreseeable future.

There are very few instances in any content area where a single term wholly captures the indexing of concept. For example, if one is interested in administrators, then a quality search would search for administrators OR narrower terms such as principals, coordinators, superintendents. The typical user used one OR in every other search and performed two to three queries per search session. In contrast, the experts used six times as many ORs and typically conducted twice as many searches. The results for the non-expert groups is quite disappointing. Most patron searches cannot

BEST COPY AVAILABLE

possibly capture subject matter nuances.

The search engine at ERICAE incorporates several recent advances from information science. The more-like function allows patrons to take descriptors from a relevant citation and recycle them into a new search. Only a handful of people of the 27,000 people searching ERIC from the ericae.net web site each month take advantage of that feature. Another advanced feature, concept searching, allows the user to automatically load a term and its narrower terms into a query. Again, only a handful of people take advantage of that option. Only about one-third of patrons are using the on-line ERIC thesaurus to help craft their queries. Not using the ERIC thesaurus is the same as guessing which terms were used by the ERIC indexers. Thus, not only is the typical end-user doing a poor job of searching, they are not taking advantage of the available tools.

It appears that, when searching the ERIC database on-line, users are satisfied if they find anything that is relevant. Their expectations appear to be low and they appear to be easily pleased. This does not bode well for the quality of the resulting research or policy decisions. The data imply that educational research and practice is not building on what has already been learned. As more end-users search for themselves, will we witness a decline in quality?

On the bright side, one in ten patrons noted that, rather than searching the literature themselves, they would prefer to have an information professional search for them. As shown in Table 4, sizeable percentages of K-12 teachers, K-12 staff, and parents value expert help. It appears that quality reference service assistance, such as the type of help that was available 15 years ago, is still valued by many. However, the vast majority of key patron groups, K-12 administrators and college professors, prefer to search for themselves. I suspect they do not realize how ineffectively they are searching.

Table 4: Searching preferences by user group.

	Do you prefer to:			
	Search for		Have a	
	Row %	Count	Row %	Count
K-12 Teacher	87.4%	560	12.6%	81
K-12 Staff	72.7%	56	27.3%	21

K-12 Staff	72.7%	56	27.3%	21
K-12 Administrator	93.4%	113	6.6%	8
College Professor	93.8%	196	6.2%	13
Parent	84.7%	61	15.3%	11
Researcher	88.8%	395	11.2%	50
Other	93.2%	384	6.8%	28
UG Student	88.4%	336	11.6%	44
Graduate Student	88.8%	796	11.2%	100
All Users	89.1%	2897	10.9%	356

On the negative side, it appears that demand for professional help is being met by non-experts. We asked patrons how often they searched for others. As shown in Table 5, almost half (47%) of the non-librarians said they occasionally or often search for others. A check on the quality of searches for those that never search for others and those that do revealed no meaningful differences in terms of number of ORs, time searching, use of the thesaurus, or hits examined. Further, there are no meaningful differences in search quality between those who report they have minimal database experience and those who occasionally search for others. Most nonprofessionals searching for others are not doing any better than are inexperienced people who search for themselves.

Table 5: Frequency of Searching for Others						
	How often do you search for others?					
	Never		Occasionally		Almost always	
	Count	Row %	Count	Row %	Count	Row %

Capacity	**K-12 Teacher**	377	58.8%	260	40.6%	4	.6%
	K-12 Staff	17	22.1%	55	71.4%	5	6.5%
	K-12 Administrator	36	29.8%	82	67.8%	3	2.5%
	College Professor	87	41.6%	118	56.5%	4	1.9%
	Parent	33	45.8%	38	52.8%	1	1.4%
	Researcher	213	47.9%	203	45.6%	29	6.5%
	Other	200	48.5%	183	44.4%	29	7.0%
	UG Student	231	60.8%	146	38.4%	3	.8%
	Graduate Student	540	60.3%	341	38.1%	15	1.7%
	TOTAL Non-Librarians	**1734**	**53.3%**	**1426**	**43.8%**	**93**	**2.9%**
	K-12 Librarian	6	8.5%	56	78.9%	9	12.7%
	College Librarian	11	11.5%	41	42.7%	44	45.8%
	TOTAL Librarians	**17**	**10.2%**	**97**	**58.1%**	**53**	**31.7%**

Thus, based on this data, it appears that

1. End-users are not doing a very good job searching on-line
2. Most end-users prefer to search for themselves
3. Many unqualified end-users are conducting searches for others who want search assistance.

These finding are consistent with the large body of pre-Internet literature and the emerging Internet era literature claiming that most end-users do obtain poor results when searching for themselves (Lancaster, Elzy, Zeter, Metzler and Yuen, 1994; Bates, Siegfried and Wilde, 1993; Tolle and Hah,

1985; Teitelbaum and Sewell, 1986). Researchers comparing faculty and student searches of ERIC on CD-ROM to searches conducted by librarians, (Lancaster, Elzy, Zeter, Metzler and Yuen 1994), for example, noted that most of the end-users found only a third of the relevant articles than were found by the librarians. With regard to web searching, Nims and Rich (Nims and Rich, 1998) studied more than 1,000 searches conducted at Magellan and noted only 13 percent of the searchers used any Boolean operators.

Perhaps now more than ever, there is a need to train end-users. Teaching search skills should be part of every introduction to research course, and searching should be taught by trained reference professionals. Training should go well beyond the traditional use of boolean logic to include sound search strategies such as expanding the query by ORing appropriate narrower and related terms, using a thesaurus to find useful descriptors, using building block or pearl building methods, and conducting multiple searches.

Where reference services are available, they should be promoted. Where they don't exit, they should be provided. In the medical field, for example, it is still common for highly qualified reference personnel to conduct searches.

I have to wonder whether we have highly qualified, well-supported reference personnel serving the K-12 community. First, why were these people searching the ERIC database at ericae.net? The CD-ROM products have a much better interface and allow for better searching. Second, have they been adequately trained in reference services? The quality of their searches were not much better than those of non-professional novices.

There is a large and growing body of literature recognizing the need for expanded reference services in today's information rich world (e.g., Blair, 1992; Buckland, 1992). While much of the literature appears to focus on training reference professionals, others proposed using software and electronic content to emulate interaction between the reference librarian and the library patron (Crane, 1992). Popular lines of research in information retrieval today include natural language processing, search engines that incorporate artificial intelligence, probabilistic logic, query by example, query expansion, automatic summaries, and concept-based searching (Lager, 1996). While tools that have resulted from these lines of research have great potential, their power cannot be realized with simple one or two word searches. The ericae.net site offers several advanced searching features (natural language processing, query by example, concept-based searching), yet they are rarely used by most end-users.

Today's attention to database creation and better search engines fails to

address a critical consumer need. Better digital libraries and more powerful search engines will not get quality materials into the hands of the end-user. Developers of digital libraries must work with content experts to develop an array of information products that help users identify and understand the available resources. These products might:

- include an introduction to the topic prepared by a key researcher in the field,
- outline issues,
- identify the most respected citations on all sides of the issue,
- contain dynamic, fully-formed, searches of the digital library, and
- identify relevant internet resources.

It would be good to have subject matter experts review resource materials, and to periodically update them. Such a resource would help ensure that novices have a better understanding of their topics and are pointed to quality references. Those wanting to conduct more in-depth examinations would have the tools and directions to do so.

References

Bates, M. J., Siegfried, S. and Wilde, D. N. (1993). An Analysis of Search Terminology Used by Humanities Scholars: The Getty Online Searching Project Number 1. Library Quarterly, 63(1), 1-39.

Blair, J. (1992). The Library in the Information Revolution. Library Administration & Management, 6 (2), 71-76 Spring 1992.

Buckland, M. (1992). Redesigning Library Services: A Manifesto. Chicago: ALA Books.

Crane, D.J. (1996). Creating Services for the Digital Library. In Online Information 96. Proceedings of the International Online Information Meeting (20th, Olympia 2, London, England, United Kingdom, December 3- 5, 1996) ERIC Document Reproduction Service No ED411861.

Evans, J. E. (1995). Economics of Information Resource Utilization: Applied Research in the Academic Community. ERIC Document Reproduction Service No ED380107.

Hertzberg, S. and Rudner, L (1999). The Quality of Researchers' Searches of the ERIC Database. Education Analysis Policy Archives, 7(25). Available online: <http://epaa.asu.edu/epaa/v7n25.html>.

Lancaster, F. W., Elzy, C., Zeter, M. J., Metzler, L. and Yuen, M. L.. (1994). Comparison of the Results of End User Searching with Results of

Two Searching by Skilled Intermediaries. RQ, 33(3), 370-387.

Lager (1996). Spinning a Web Search. Available online <http://www.library.ucsb.edu/untangle/lager.html>.

Nims, M. and Rich, L. (March 1998). How Successfully Do Users Search the Web. College and Research Libraries News. 155-158.

Tolle, J. E. and Hah, S. (1985). Online Search Patterns. Journal of American Society of Information Science, 36(2), 82-93.

Teitelbaum, S. and Sewell, W. (1986). Online Searching Behavior Over Eleven Years. Journal of American Society of Information Science, 37(4), 234-245.

Copyright © 2000 Lawrence Rudner